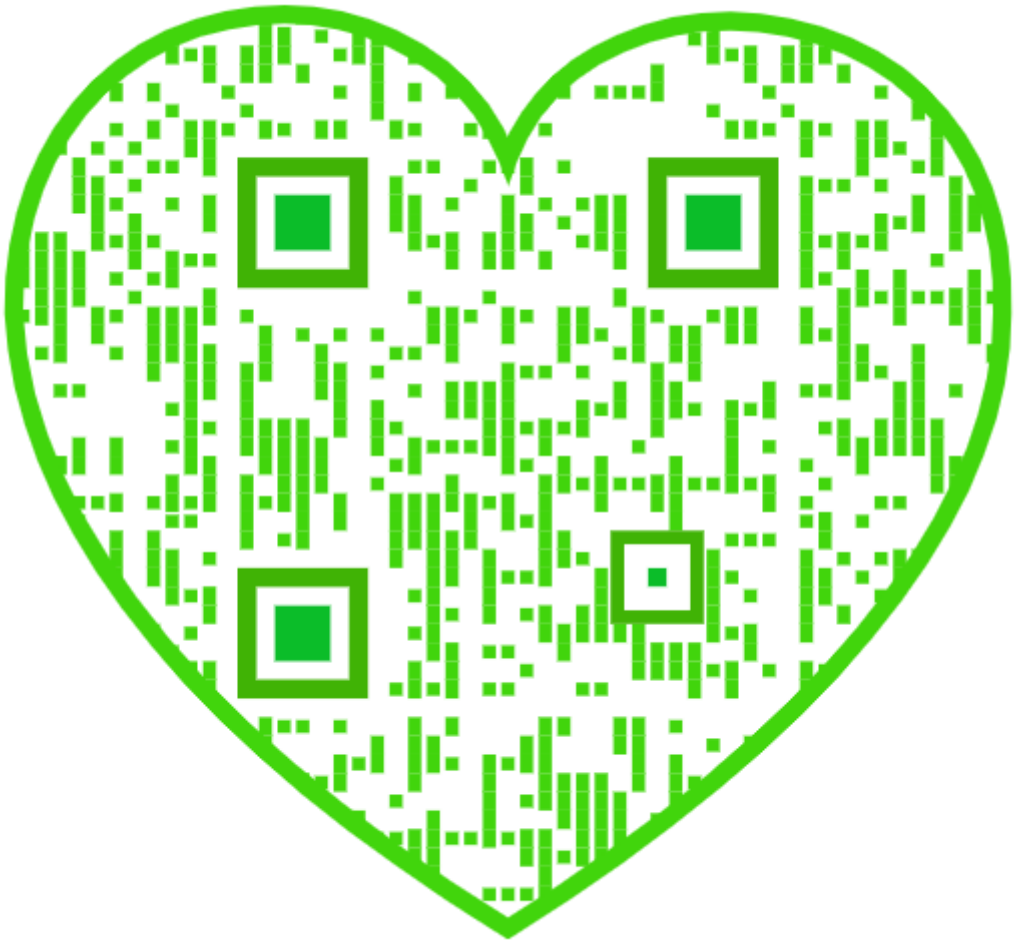
Master in Artificial Intelligence

Algorithm Selection & Development XXI



Purpose

The purpose of the section is to help you learn how to research, select, and develop appropriate algorithms to become a Successful Artificial Intelligence (AI) Engineer

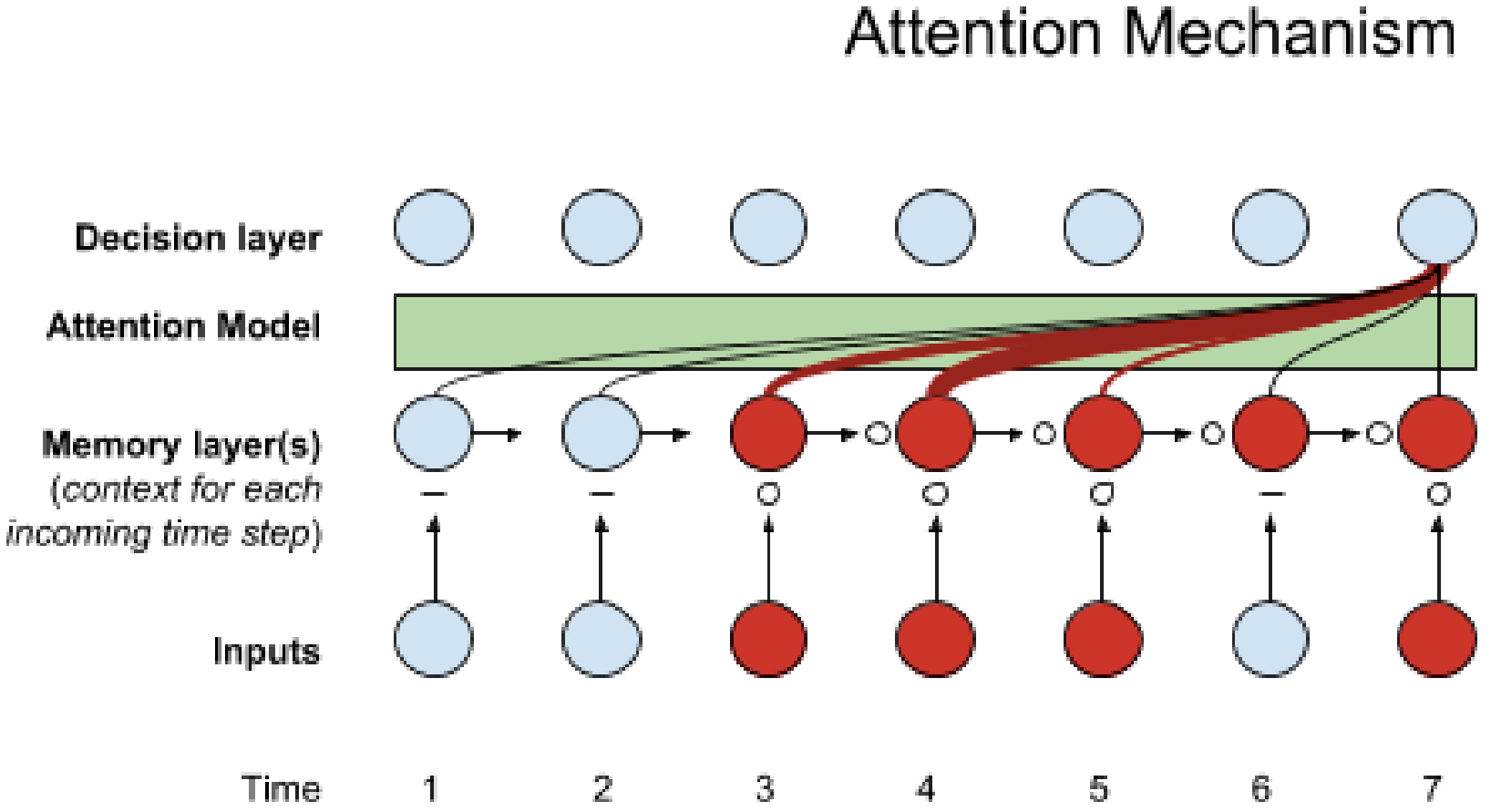
At the end of this lecture, you will learn the following

- **What are attention mechanisms in deep learning models to consider the interpretability and explainability of the selected models**

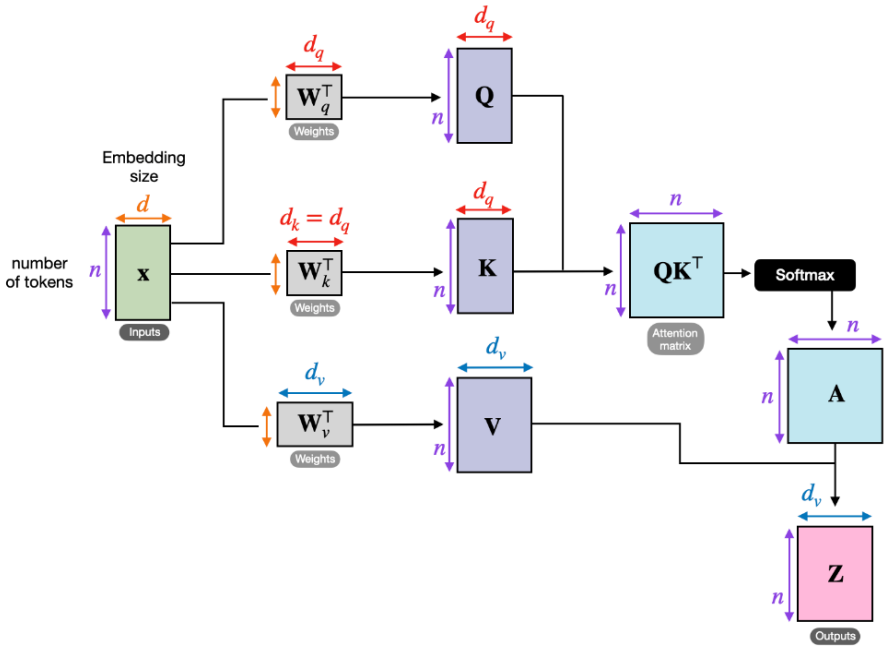
What are attention mechanisms in deep learning models to consider the interpretability and explainability of the selected models

Interpretability and Explainability techniques

- Feature importance analysis
- Model interpretability methods
  - SHAP values
  - LIME
- Attention mechanisms in deep learning models


Attention Mechanism
Decision layer
Attention Model
Memory layer(s)
(context for each
incoming time step)
Inputs
Time
1
2
3
4
5
6
7

Self-Attention Mechanism



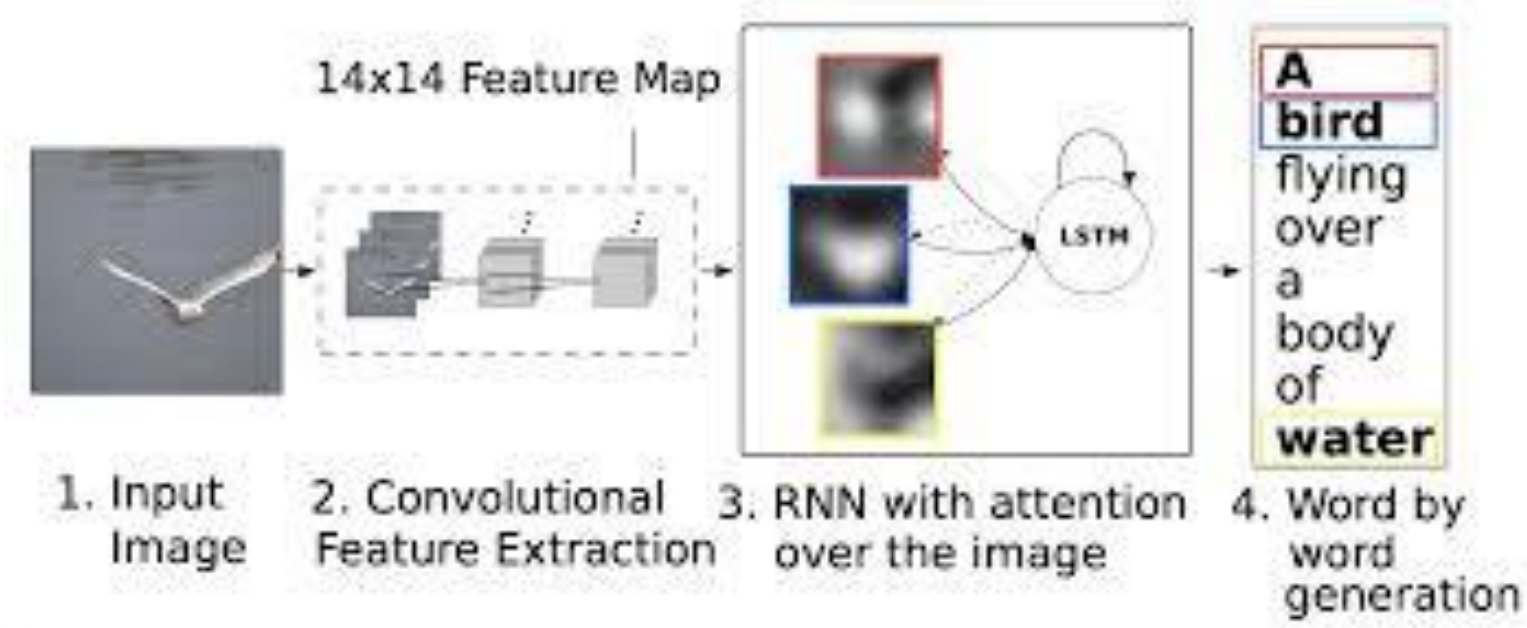
Interpretability

- Visualizing attention weights, one can see which parts of the input sequence are attended to more heavily during the prediction process
- Providing insights into the model's decision-making process.

Explainability

- Attention weights can be used to explain model predictions
- Highlighting the words or tokens that contribute most to the output

Visual Attention Mechanism



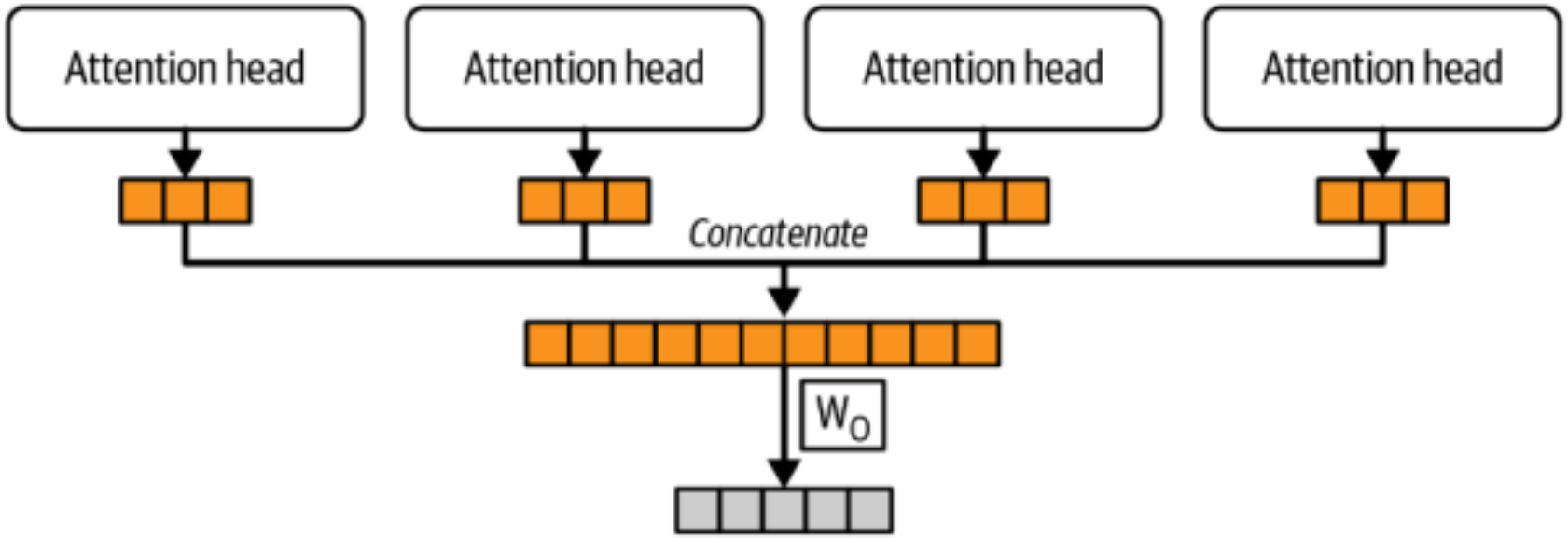
Interpretability

- Visualizing the attention maps can reveal which regions of the input image are important for the model's predictions
- Helping users understand where the model is looking and what features it is focusing on

Explainability

- Attention maps provide a visual explanation of the model's decision-making process by
- Highlighting the salient regions of the input image that influence the output

Multi-Head Attention



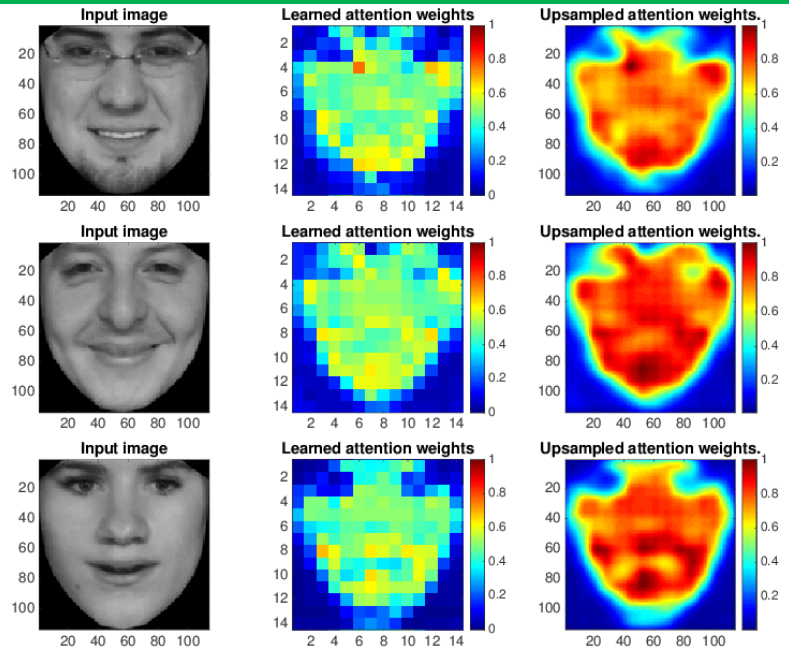
Interpretability:

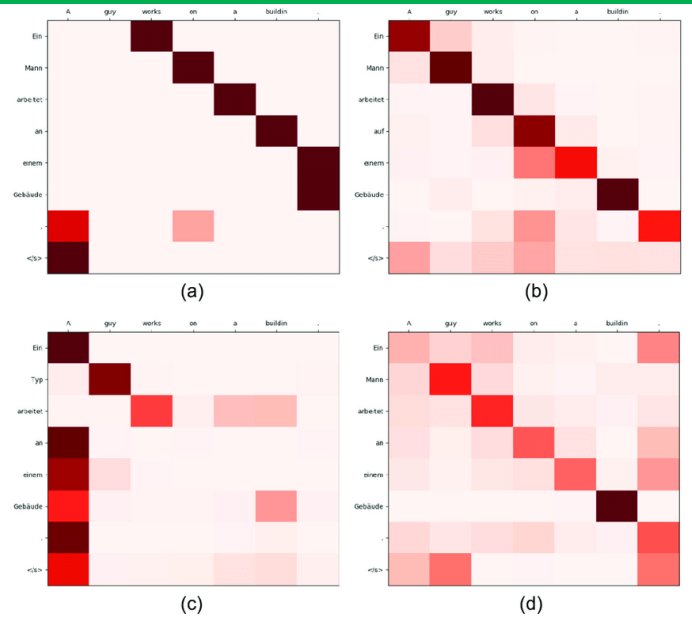
- Can gain a more comprehensive understanding
- How different parts of the input interact and influence the model's predictions.

Explainability:

- Capturing various aspects of the input data
- Their contributions to the output

Attention Visualization Techniques





SpanBert - SQuAD

Question: who is not an entrepreneur ?

Context: steven is an author . annie is an entrepreneur .

Answer: Annie

Interpretability

- Provide intuitive and interpretable representations of attention weights
- Making it easier to understand how the model processes input data.

Explainability

- Can gain insights into the model's decision-making process
- Understand which features or components are most influential for specific predictions

What are attention mechanisms in deep learning models to consider the interpretability and explainability of the selected models

Interpretability and Explainability techniques

- Feature importance analysis
- Model interpretability methods
  - SHAP values
  - LIME
- Attention mechanisms in deep learning models

What is next?

How to research, select, and develop appropriate algorithms

Deployment and Monitoring

- Deploy the trained model in a production environment and integrate it into the application workflow
- Implement monitoring and logging mechanisms to track model performance, drift, and errors over time
- Continuously evaluate and update the model as new data becomes available or the problem requirements change

Master in Artificial Intelligence

Algorithm Selection & Development XXI